Big Data Analytics with IBM Cognos Dynamic Cubes

IBM Redbooks Solution Guide

IBM® Cognos® Dynamic Cubes, which is a feature of the Cognos Business Intelligence V10.2 software, complements the existing query engine. It extends Cognos scalability to enable speed-of-thought analytics over terabytes of enterprise data, without being forced to rely on a new data-warehousing appliance. With this capability, which adds a level of query intelligence, you can unleash the power of your large enterprise data warehouse. Figure 1 illustrates how the Cognos Dynamic Cubes feature is integrated into the Cognos Business Intelligence stack.

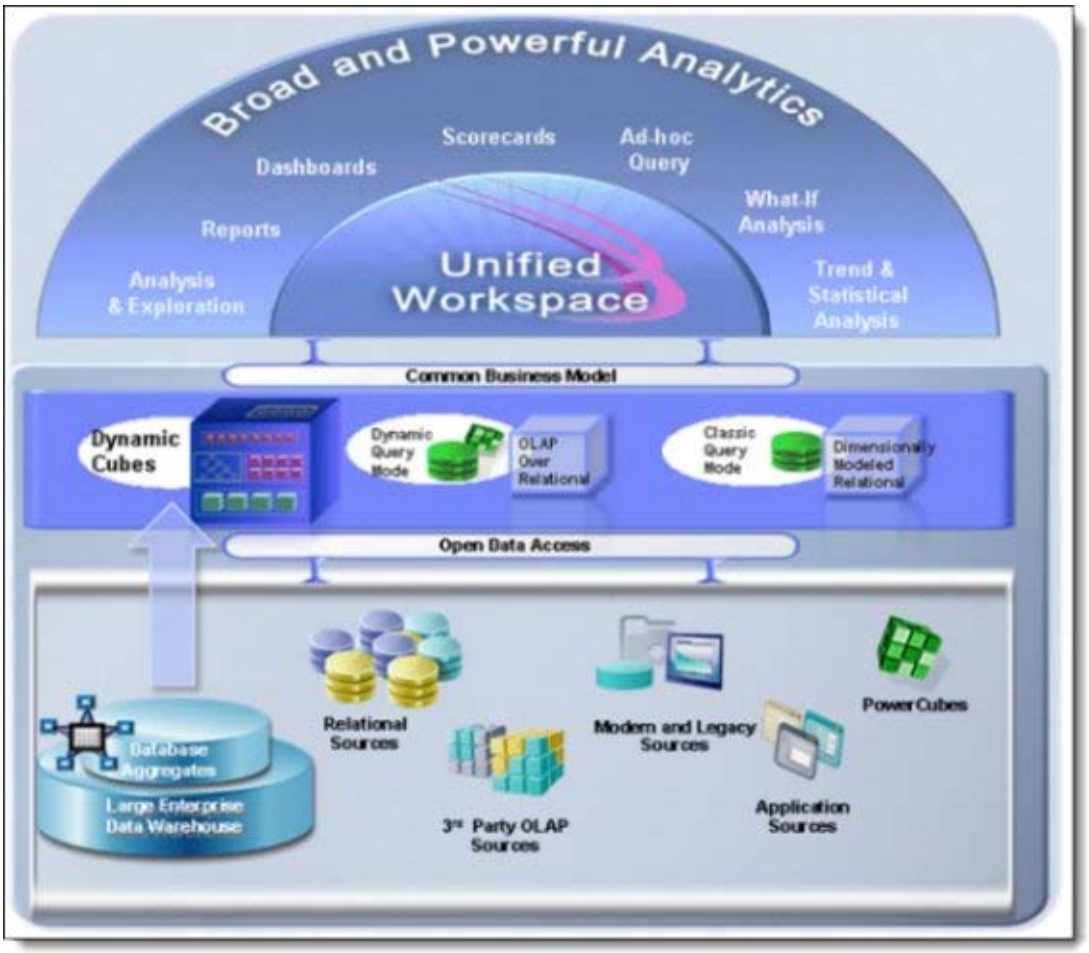

Figure 1. Cognos Dynamic Cubes integrated into the Cognos Business Intelligence stack

Did you know?

In 2012, about 2.5 quintillion bytes of data were created every day. Social media is now generating petabytes of data per day.

Business value

With social data generating petabytes per day, and instrumented devices becoming the norm, data volume growth is accelerating at an unprecedented pace. Big data is a growing business trend, with data from unconventional sources having the potential to be business disruptors. However, before the power of these new sources can be fully used, you must understand what is happening within your own business. Understanding your own business is added value of a data warehouse and is why taking full advantage of these data holdings is a critical first step to using these new sources of data. In addition, any organization that relies on instrumented infrastructures can maximize the efficiency of its operations. Analytics is key to accomplishing this type of optimization, leading to concrete business results.

Data warehouses are the recognized foundation for enterprise analytics. By using data warehouses, an organization can bring together cleansed data from separate sources of input, both internal and external, such as from partners or suppliers. Instead of garbage in, garbage out information to support decision-making, a consistent and consolidated enterprise-wide view of data from a business provides the foundation to improve your business. Building upon a trusted information platform for analytics is a key contributor to long-term business health. Not only do data warehouses enable higher quality information, they enable high-performance data access for analytic-style applications. IBM Cognos Dynamic Cubes technology helps to use the core strengths of an enterprise data warehouse and take it to the next level of performance for analytics, making the deploying and tuning easier and faster.

Solution overview

The IBM Cognos Dynamic Cubes technology is meant to solve a specific but growing business problem, enabling high-performance interactive analysis over terabytes of data in an enterprise data warehouse. As data volumes grow, analyzing that data with speed-of-thought performance can be challenging. Even with modern data warehouse technology, some operations require significant computation or data movement. This computation or movement creates delays and reduces the satisfaction of business users who want to perform these analyses.

Various ways exist to accomplish performance over large volumes of data. From self-contained cubes to large in-memory appliances, different vendors are employing variations of similar methodologies to give business users timely response times. The Cognos Dynamic Cubes technology aims to give maximum flexibility in how memory is used to accelerate interactive analysis over terabytes of data so that you can evolve your deployments over time.

Solution architecture

The IBM Cognos Dynamic Cubes technology is part of the IBM Cognos Business Intelligence query stack and is available with existing IBM Cognos entitlements. It provides a new and powerful tool to enable high performance analytics over large data warehouses (see Figure 2).

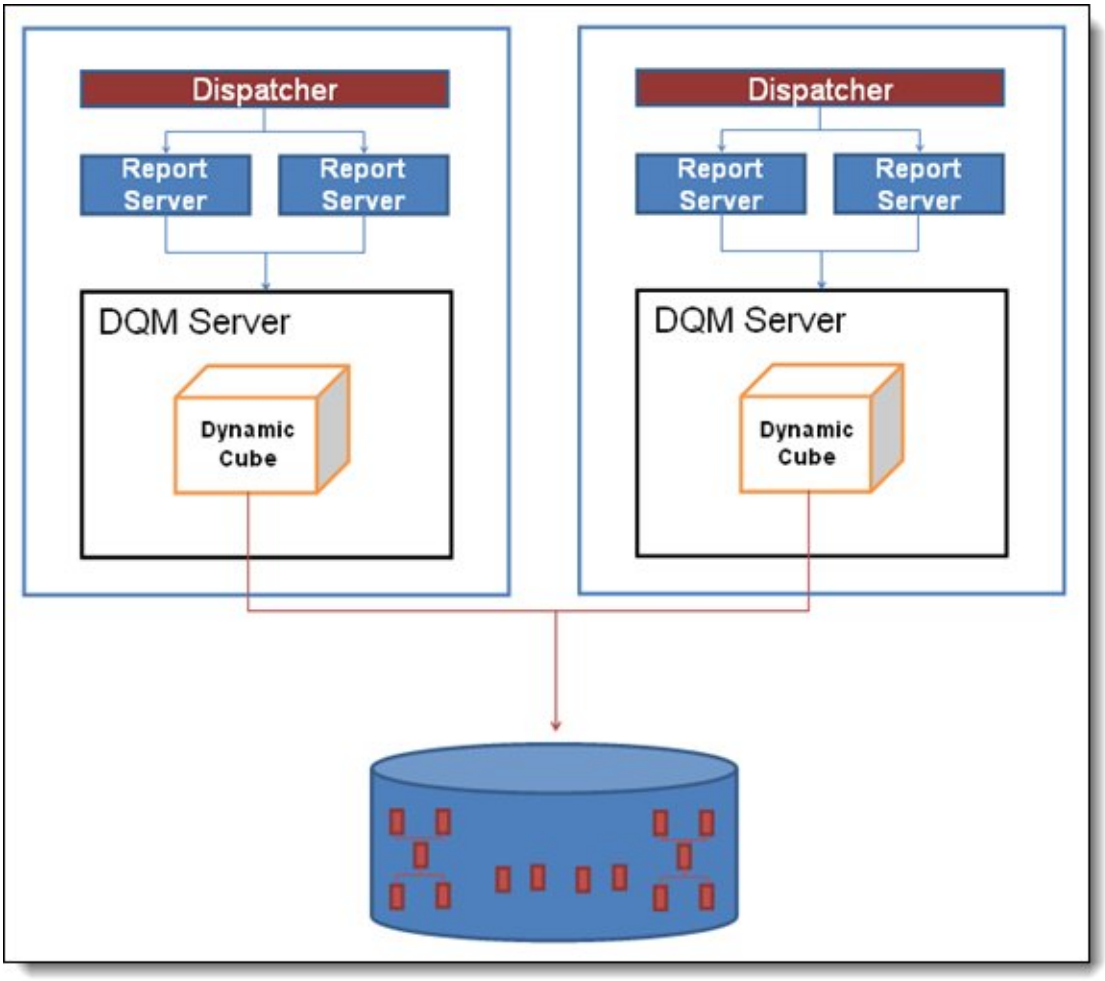

Figure 2. A high-level representation of the Cognos Dynamic Cubes architecture

The Cognos Dynamic Cubes solution consists of IBM Cognos Cube Designer (a modeling tool), a dynamic cube object in the administration environment (which becomes the data source), and Aggregate Advisor (a wizard) that is started from Dynamic Query Analyzer:

- IBM Cognos Cube Designer

  Cognos Cube Designer is a modeling tool that brings together the best modeling principles from past successful modeling technology, with a modern and extensible architecture. The first step to deploying Cognos Dynamic Cubes is to model with the Cognos Cube Designer.

- IBM Cognos Dynamic Cubes Server

  After a dynamic cube is designed and deployed, it becomes available in the Cognos environment and acts as the data source to the interface layer for dynamic cube applications. It manages all aspects of data retrieval and leverages memory to maximize responsiveness, giving you full flexibility to manage what is in memory and when you want to refresh in-memory data. You manage dynamic cubes in the Cognos administration environment.

  A dynamic cube contains several in-memory elements to drive performance:

  - Metadata members
  - Aggregates
  - Data
  - Results sets
  - Expressions

- Aggregate Advisor (part of IBM Cognos Dynamic Query Analyzer)

  The Aggregate Advisor scans cube definitions and usage logs. It also recommends aggregates to improve performance. This approach helps to more easily address specific performance problems.

Usage scenarios

The Cognos Dynamic Cubes solution applies to the following usage scenarios.

Multigrain fact scenarios

A common modeling problem is multigranularity, which occurs when levels of dimensional detail for facts differ. In your data source, the level of information in a dimension can be more precise than the fact data of some fact tables in which the dimension takes part. For example, a time dimension can have dimension information for the levels of year, quarter, month, and day. For a sales fact table, the facts exist at the day level. For a fact table with planned sales values, the fact grain or level is probably at a higher level of detail, such as month. This difference in fact grain can make it difficult to plan queries correctly if a report user included a level below the fact grain in the report.

The method for handling this scenario is to create role-playing dimensions for each instance of a dimension that has varying levels of granularity to different fact tables. In this example, you need to model a time dimension down to the day level for the sales cube and to model a time dimension down to the month level for the sales target cube. For the day level, the relationship is formed between the appropriate keys in the fact table and the day level. For the month level, the relationship is formed between the month level and the fact table keys.

If you want to create reports that used data from both cubes, you can create a virtual cube. The virtual cube can have the two time dimensions merged. The levels of both source cubes will be merged. The virtual cube might have levels that are below the grain of measures from one of the source cubes. Queries that are made in the virtual cube that use dimension levels below the grain of a measure will return null values, ensuring that the consumers of the cube do not have double counting.

Perhaps you have the following cubes:

- A sales fact cube, where the fact grain for the time dimension is at the day level
- A planned sales cube, where the fact grain for the time dimension is at the month level

By using a virtual cube that employs both of these cubes as its source, you can make a query with day-level objects (or a member of that level, depending on the studio that you are using) against the sales facts. Then, you get the results and the expected null values for the planned sales facts. If you used a time dimension grain that was common to both fact tables, you get non-null values for measures from both fact tables.

Aggregate cubes

A second scenario addresses situations where a higher level of granularity is desirable through aggregate awareness. This awareness is accomplished through aggregate cubes, which are supported by IBM Cognos Dynamic Cubes. *Aggregate cubes* define the measures, dimensions, and dimension grain by which queries can be routed to aggregate tables rather than to the detail fact table. Because aggregate tables store fact data at a higher-than-detail level of granularity, the time necessary to aggregate values during the query can be lessened, improving performance. A query can be routed to the aggregate table if all the measures and dimension hierarchies of the query exist in the aggregate cube definition. Not all of the dimensions and measures in the aggregate cube need to be in the query.

The objective of modeling an aggregate cube is to establish rules by which the dynamic cube can detect when it can route a query to an aggregate table. This task is done by specifying a mapping from the identifiers in the dimensions and measures in the cube that have scope to the aggregate table, to the identifiers in the aggregate table. If necessary, mapping is done to its related tables in a rolled-up dimension schema.

This aggregate cube routing directs a query only to the aggregate table for a query that uses objects from a dimension grain at or above the grain of the mapping between it and the aggregate table. Therefore, using objects from a grain below the mapping grain does not cause double-counting because that query continues to route to the detail fact table.

Integration

IBM Cognos Dynamic Cubes is tightly integrated into the Cognos Business Intelligence stack, and its data can be surfaced through any of the Cognos interfaces. With this method, existing customers can integrate this technology into their application environment without affecting existing users. Such users are already familiar with interfaces such as Report Studio, Business Workspace, and Business Workspace Advanced (previously named Business Insight and Business Insight Advanced).

Different data requirements require different data solutions. One data path cannot be proficient at solving widely different data problems. Therefore, IBM Cognos has technologies that are built to suit specific application requirements. Table 1 can help you better understand the primary use case for each technology. However, carefully consider your individual application requirements when you make such a decision.

Table 1. Use cases for Cognos Dynamic Cubes technology

Cube technology	Primary use cases
IBM Cognos TM1®, in-memory cube technology with write-back support	• It is optimal for write-back, what-if analysis, planning and budgeting, or other specialized applications. • It can handle medium data volumes. The cube is run 100% in memory. • Aggregation occurs on demand, which can affect performance with high data and high user volumes.
Dynamic Cubes, in-memory accelerator for dimensional analysis	• It is optimal for read-only reporting and analytics over large data volumes. • It provides extensive in-memory caching for performance, backed by aggregate awareness to use the power and scalability of a relational database. • A star or snowflake schema is required in the underlying database (used to maximize performance).
PowerCubes, file-based cube with pre-aggregation	• It is optimal to provide consistent interactive analysis experience to many users when the data source is an operational or transactional system, and a star or snowflake data structure cannot be achieved. • The pre-aggregated cube architecture requires careful management, by using cube groups to achieve scalability. • Data latency is inherent with pre-aggregated cube technology, where data movement into the cube is required.
OLAP Over Relational (OOR), dimensional view of a relational database	• It is optimal to easily create a dimensional data exploration experience over low data volumes in an operational or transactional system, and where latency must be carefully managed. • Caching on the Dynamic Query server helps performance. • Processing that is associated with operational or transactional system affects performance.

Supported platforms

For information about software environments that are supported in IBM Cognos Business Intelligence V10.2, see "Cognos BI 10.2 Software Environments" at:
http://www.ibm.com/support/docview.wss?uid=swg27027080

Ordering information

Table 2 lists the programs that IBM Cognos Business Intelligence contains.

Table 2. Program numbers and names

Program number	Program name
5724-W12	IBM Cognos Business Intelligence V10.2.0
5724-W68	IBM Cognos Business Intelligence PowerPlay® V10.2.0
5724-W19	IBM Cognos Business Intelligence Reporting V10.2.0
5724-W13	IBM Cognos Data Manager V10.2.0
5724-W20	IBM Cognos Mobile V10.1.0

Related information

For more information, see the following documents:

- *IBM Cognos Business Intelligence V10.1 Handbook,* SG24-7912
  http://www.redbooks.ibm.com/abstracts/sg247912.html?Open
- *IBM Cognos Business Intelligence V10.1: Intelligence Unleashed,* REDP-4693
  http://www.redbooks.ibm.com/abstracts/redp4693.html?Open
- IBM Cognos Business Intelligence 10.2.0 Information Center
  http://pic.dhe.ibm.com/infocenter/cbi/v10r2m0/index.jsp
- Dynamic Cubes Installation and Configuration Guide 10.2.0
  http://pic.dhe.ibm.com/infocenter/cbi/v10r2m0/nav/0_9
- Dynamic Cubes User Guide 10.2.0
  http://pic.dhe.ibm.com/infocenter/cbi/v10r2m0/nav/5_6
- Dynamic Query Analyzer Installation and Configuration Guide 10.2.0
  http://pic.dhe.ibm.com/infocenter/cbi/v10r2m0/nav/0_7
- Dynamic Query Analyzer User Guide 10.2.0
  http://pic.dhe.ibm.com/infocenter/cbi/v10r2m0/nav/5_8
- IBM announcement letters and sales manuals
  http://www.ibm.com/common/ssi/index.wss?request_locale=en

  On this page, enter `IBM Cognos Business Intelligence`, and click **Search**. On the next page, narrow your search results by information type, geography, language, or all three options.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service. IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you. This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document was created or updated on December 7, 2012.

Send us your comments in one of the following ways:

- Use the online **Contact us** review form found at:
  ibm.com/redbooks
- Send your comments in an e-mail to:
  redbook@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

This document is available online at http://www.ibm.com/redbooks/abstracts/tips0942.html .

Trademarks